

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Facultad de Ingeniería

Maestría en Explotación de Datos y Descubrimiento del Conocimiento
Materia: Aprendizaje Automático

Docentes a cargo:

Profesores:

Dra. Viviana Cotik: [vcotik at dc.uba.ar](mailto:vcotik@dc.uba.ar)

Msc. Guillermo Henrión

Auxiliares:

Ing. Leonardo Pepino

Lic. Gastón Bujía

Programa

Duración: un cuatrimestre

Horas de clases semanales: 3 hs entre clases teóricas y prácticas

Objetivo

El Aprendizaje Automático o Machine Learning es una rama de la Inteligencia Artificial, que estudia algoritmos para que las computadoras aprendan a resolver problemas a partir del uso de datos sin ser programadas específicamente.

Esta materia introduce las principales técnicas de Aprendizaje Automático. El objetivo de este curso es dotar a los alumnos del conocimiento de las principales técnicas, su uso correcto y del conocimiento de herramientas para la implementación de los algoritmos.

Al finalizar el curso el alumno deberá:

- conocer las principales técnicas y aplicaciones del área
- saber decidir qué técnica utilizar para resolver distintas problemáticas
- poder implementar los algoritmos estudiados
- contar con los recursos para aplicar las técnicas de manera correcta
- tener recursos para estudiar y aplicar nuevas técnicas

Contenido

El temario se realiza teniendo en cuenta que se trata de un curso de 3 horas semanales de clases teórico-prácticas.

- **Introducción al aprendizaje automático.** Definición. Aprendizaje automático supervisado, no supervisado. Aprendizaje por refuerzos.
- **Datos.** Disponibilidad. introducción a big data. Calidad de datos. Sesgo en los datos. Cuestiones éticas. Posibles Problemas en Aplicación de técnicas de Aprendizaje automático. Otras definiciones.
- **Aprendizaje de conceptos.** Aprendizaje inductivo. Adquisición de conceptos como búsqueda. Espacio de búsqueda y relación de orden General-Específico. Algoritmos FIND-S, List-then-eliminate, Eliminación de candidatos. Sesgo inductivo
- **Árboles de decisión.** Algoritmo: criterio de selección, sesgo inductivo, Occam's Razor, sobreajuste, poda. Adecuación a valores continuos. Valores faltantes. Atributos con costo.
- **Evaluación y selección de modelos.** Indicadores de performance y error. Cross-validation y métricas
- **K-Nearest neighbor**
- **Naive Bayes y redes bayesianas.**
- **Introducción a Redes neuronales artificiales:** Inspiración biológica. Modelo matemático. Funciones de activación. Arquitectura feed-forward. Perceptrón simple y perceptrón multicapa. Función de pérdida y función de costo. Método del descenso por gradiente. Forward y backward pass. Distintas arquitecturas de RN y resurgimiento de RN.
- **Ensamblados/ combinación de múltiples modelos:** Sesgo y Varianza. Bagging, Boosting, Random Forest, Stacking

Los temas serán presentados conceptualmente por los docentes y sustentados por la lectura de textos específicos y, de ser posible mediante su implementación o desarrollo. Para cada tema a tratar se promoverá la lectura de refuerzo.

Las clases suelen tener un componente teórico-conceptual (fundamentalmente dictado por el docente a cargo) y uno práctico (fundamentalmente dictado por el docente auxiliar). El material de trabajo se publica en una plataforma virtual ordenado por clase.

Modo de evaluación:

Exámen individual

Se podrá recuperar una sola vez al final del cuatrimestre.

Trabajos prácticos

Los trabajos prácticos (TPs) tendrán como objetivo afianzar los conceptos mediante la utilización de software y ejercitación. Se prevé la realización de 2 trabajos prácticos domiciliarios elaborados en grupos de exáctamente 3 alumnos. Los TPs se presentan con un informe, que debe estar redactado de forma acorde al nivel de educación de un alumno de postgrado.

Bibliografía

Básica:

- Mitchell, ["Machine Learning"](#), McGraw-Hill, 1997.
- James, Witten, Hastie & Tibshirani, ["An Introduction to Statistical Learning with Applications in R"](#), 6th ed, Springer, 2015.
- Marsland, ["Machine Learning. an Algorithm Perspective"](#), CRC Press, 2015
- Alpaydin, ["Introduction to Machine Learning"](#), 2010.
- Müller & Guido, ["Introduction to Machine Learning with Python"](#), O'Reilly, 2016.

Avanzada:

- Hastie, Tibshirani & Friedman, ["The Elements of Statistical Learning"](#), 2nd ed, Springer, 2009.
- Bishop, ["Pattern Recognition and Machine Learning"](#), Springer, 2006.
- Seni, Elder, ["Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions"](#), Morgan & Claypool, 2010.
- Leskovec, Rajaramán, Ullman, ["Mining of Massive Datasets"](#), 2010.